Speedup of GEMM on A100 and RTX 4090 (FP16)